Published in final edited form as:
Epidemiology. 2011 July ; 22(4): 599–606. doi:10.1097/EDE.0b013e31821d0879.

Left Truncation, Susceptibility, and Bias in Occupational Cohort Studies

Katie M. Applebaum[a,b], **Elizabeth J. Malloy**[c], and **Ellen A. Eisen**[b,d]

[a]Department of Epidemiology, Boston University School of Public Health, Boston, MA

[b]Department of Environmental Health, Harvard School of Public Health, Boston, MA

[c]Department of Mathematics and Statistics, American University, Washington, DC

[d]Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, CA

Abstract

Background—Left truncation occurs when subjects who otherwise meet entry criteria do not remain observable for a later start of follow-up. We investigated left truncation in occupational studies due to inclusion of workers hired before the start of follow-up in a simulation study.

Methods—Using Monte Carlo methods, we simulated null and positive associations between exposure (work duration) and mortality for 500 datasets of 5000 subjects, assuming the absence and presence of heterogeneity in susceptibility to disease and to the effect of exposure. We examined incident hires (followed since hire) and left-truncated prevalent hires (those hired before baseline and remained employed at baseline). We estimated the association ($\hat{\beta}_1^*$) as the mean slope using Cox proportional hazards with a linear term for exposure, under scenarios with and without susceptibility.

Results—With homogeneous susceptibility, there were no differences between incident and prevalent hires. Introducing only disease susceptibility did not change results. However, with heterogeneous susceptibility to the effect of exposure, downward bias was observed among prevalent hires under both the true null and positive exposure-response scenarios. The bias increased with time between hire and baseline (null: $\hat{\beta}_1^*=0.05\,[\mathrm{SD}=0.08]$, $\hat{\beta}_1^*=-0.08\,[\mathrm{SD}=0.24]$, $\hat{\beta}_1^*=-0.18\,[\mathrm{SD}=0.98]$ if hired <15, 15 to <30, and ≥30 years before baseline, respectively), coincident with a decreasing percentage of susceptible subjects.

Conclusions: Prevalent hires induce downward bias in an occupational cohort. This occurs because subjects who are less susceptible to the exposure remain exposed the longest, thereby underestimating the association.

Correspondence: Katie M. Applebaum, Department of Epidemiology, Boston University School of Public Health, 715 Albany St, T322E, Boston, MA 02118. kappleba@bu.edu.

Left truncation occurs in a cohort when subjects at risk prior to baseline do not remain observable until the start of follow-up.[1] Similar to other threats to study validity, this potential bias may lead to incorrect conclusions about the presence of an association, as well as its direction and magnitude. Left truncation has been discussed in various branches of epidemiology, including studies of time to development of AIDS among HIV patients,[2–4] the occurrence of spontaneous abortion or birth defects,[5,6] and cancer survivors interviewed some time after diagnosis.[7] Our focus is the impact of left truncation on exposure-response relationships in occupational epidemiology.

In occupational research, left truncation occurs when workers hired prior to the start of follow-up (baseline) and still working at baseline are followed over time. The terms cross-sectional-follow-up[8] or cross-sectional cohort[9] have been used to describe a cohort that consists entirely of workers identified at one point in time (eg, the start of follow-up) and followed forward. Because traditional cross-sectional studies (without follow-up) have been found to over-represent long-term, healthier workers and underestimate disease prevalence,[10] results from such studies are generally interpreted with caution. Yet, the literature offers little guidance on how to interpret follow-up studies in which workers were identified in a cross-sectional manner. Data collected during a follow-up period might be assumed to confer the full benefits of a cohort study. However, the manner in which subjects were identified, even in a follow-up study, may influence the findings.

Occupational cohort studies often comprise a blend of “prevalent hires” (eg, the cross-sectional cohort) and “incident hires.” The phrase incident hires indicates that workers were hired during the follow-up period.[11] Prevalent hires refers to those with a date of hire that precedes study baseline and who are still working at the start of follow-up. The primary concern about prevalent hires is that they may represent a healthy subset of all persons who worked prior to follow-up. Mortality rates have been found to be lower among active workers than workers who have left employment.[12–15] Likewise, prevalent hires have been found to have lower mortality than incident hires, although such contrasts are rarely reported.[8,9] Limited data suggest that, on average, prevalent hires have higher cumulative exposure, are employed longer, and are younger at hire and older at death than incident hires.[11,16] Lower relative risks at a given exposure level have been reported for prevalent hires compared with incident hires in studies of chloromethyl ethers and respiratory cancer,[16] silica and lung cancer,[11] and cadmium and lung cancer.[17] These findings suggest that the distinction between prevalent and incident hires is nontrivial.

At the core of concerns about prevalent hires in occupational cohorts is the idea of exchangeability.[18] A worker population is a mixture of subjects with varying degrees of susceptibility to the effect of exposure; exposure may have no effect on disease risk for some while cause disease in others. Assume that, by design, subjects were enrolled in a study only after they had survived for an extended period of time. If that time window was sufficiently long to allow disease to occur before the start of follow-up, then subjects who made it into follow-up may not be exchangeable with those who did not. Therefore, one of the questions we examine in this simulation study is whether inherent variability in susceptibility influences differences observed between incident and prevalent hires.

We initially examined the issue of left-truncated occupational data in a cohort of Vermont granite workers.[11] We suspected that left truncation was operating in conjunction with the healthy worker survivor effect, a bias arising when healthier workers accrue more exposure than less healthy workers who take time off work, transfer to jobs with lower exposure, or terminate employment.[10] In the present simulation study, we focus solely on the impact of left-truncated data and investigate its potential for bias in the absence of the healthy worker survivor effect. We examine bias from left truncation in a simulation study of occupational cohorts by allowing some subjects to have worked before the start of follow-up and assuming heterogeneity with respect to susceptibility (to disease, to the effect of exposure, or both). We compare exposure-response results for incident and prevalent hires for null and positive exposure-disease associations.

METHODS

Generating the Data

We used a Monte Carlo simulation method based on one previously used for occupational cohorts.[12] We generated 500 datasets of cohorts containing 5000 subjects each (using R 2.8.1, www.r-project.org). Subjects were randomly assigned an age and date of entry into the workforce. Age at entry was based on an exponential distribution with a mean of 27 years (minimum, 18 years), and date of hire was uniform across a 100-year hire period (1900–2000). The subjects were also assigned a maximum possible employment duration based on an exponential distribution with mean of 48 years. Subjects (as incident hires) were followed from time of entry in the workforce until 2050. A subject's work tenure would be reduced from the maximum possible duration for the following reasons: end of follow-up, death from competing risk or disease of interest, or employment after age 65.

To determine the occurrence of outcomes during follow-up, we generated the probability of death from the cause of interest using a logistic function that depended on age, exposure (using employment duration), and susceptibility, as follows:

$$\begin{aligned}\text{logit } P = {} & \beta_0 + \beta_1(\ln \text{work duration}) + \beta_2(\ln \text{age}) \\ & + \beta_3(\text{disease susceptibility}) \\ & + \beta_4(\text{disease susceptibility}^{*}\ln \text{work duration}),\end{aligned} \qquad \text{Model (1)}$$

where P was the probability of death from the cause of interest estimated in each person-year. The exposure-response parameter, β_1, was specified for a null ($\beta_1 = 0$) and a positive ($\beta_1 = 0.5$) association, chosen to simulate a relative risk typically observed in occupational studies. The background mortality rate was selected to yield enough cases to provide stable parameter estimates ($\alpha_0 = -0.5$ for null association, -0.7 for positive association). We also varied the association with age ($\beta_2 = 0, 1, 2$) to examine the simulation under different conditions. The mean estimates from the simulation were reproduced under these various simulation conditions; thus only one age scenario for a null and one for a positive association will be presented (null ($\beta_1 = 0$): $\beta_0 = -5$, $\beta_2 = 2$; positive ($\beta_1 = 0.5$): $\beta_0 = -7$, $\beta_2 = 1$).

To introduce heterogeneity in disease susceptibility, we defined a binary term (β_3) to indicate susceptibility to disease. We also allowed for increased susceptibility in the presence of exposure, which we will refer to as susceptibility to the effect of the exposure, by adding an interaction term between exposure and disease susceptibility (β_4). We generated data without any variability in susceptibility ($\beta_3 = 0$, $\beta_4 = 0$). We also simulated data with heterogeneity in disease susceptibility only ($\beta_3 = 0.5$, $\beta_4 = 0$), susceptibility only to the effect of exposure ($\beta_3 = 0$, $\beta_4 = 0.5$), and both ($\beta_3 = 0.5$, $\beta_4 = 0.5$). For simulations with susceptibility, we considered 2 scenarios: 25% of study subjects susceptible and 50% of study subjects susceptible.

For each subject-year, we sampled from a binomial distribution with parameter *P*, to determine whether a subject died of the cause of interest in that year. Similarly, we allowed subjects to die of competing risks, which depended only on age (parameters for background rate = −5.09, and age = 0.09, with age centered at 60 years). The length of follow-up for each subject was the time from date of hire until death from cause of interest, death from competing risk, or end of follow-up, whichever occurred first.

Modeling the Dose Response

The simulated data were modeled using Cox proportional hazards regression to examine the dose response between exposure duration and mortality due to the disease of interest. The Cox proportional hazards model is asymptotically equivalent to the pooled logistic regression used to simulate the data because this pooled logistic regression uses person-years as the unit of analysis, making the outcomes rare.[19] In the Cox models, age served as the time metric. A nested case-control sample of 20 controls per case matched on age was also used to speed the computations. Hazard ratios (HRs) were modeled using a linear term for the log of work duration as follows:

$$\ln(\mathrm{HR}) = \beta_1(\ln \text{work duration}) \quad \text{Model (2)}$$

In modeling the dose response between the log of work duration and disease over 500 simulations, we estimate $\hat{\beta}_1^*$ and expect that it will be an unbiased estimate of β_1 in model (1) when there is no heterogeneity in susceptibility ($\beta_3 = \beta_4 = 0$). For models with heterogeneity, the estimate of the effect of work duration $\hat{\beta}_1^*$ will not be an estimate of β_1, but rather a weighted estimate of the association between the log of work duration and disease among subgroups with and without susceptibility. Computationally, $\hat{\beta}_1^*$ will be an estimate of the average of β_1 and β_4 from model (1), weighted by the distribution of the person-time contributed by susceptible and nonsusceptible subjects.

Inducing Left Truncation

The simulated data represent all workers hired any time between 1900 and 2000 and followed up for mortality from 1900 to 2050. All of these subjects meet the definition of incident hires because they were hired during follow-up, not before (eg, no left truncation).

To create left-truncated data and examine the association among prevalent hires, we moved the start of follow-up to a later time point, that is, 1950. By doing so, we can identify prevalent hires as those who started working between 1900 and 1949, up to 50 years prior to the start of follow-up, and were still working when follow-up began. We examined the impact of left truncation (with and without susceptibility types) by comparing mean cohort characteristics and dose-response estimates ($\hat{\beta}_1^*$) across the 500 cohorts for the incident hires and the prevalent hires.

We further examined the impact of left truncation among prevalent hires by investigating the relationship between bias and the length of time between hire date and baseline in 3 strata: hired ≥30 years before baseline, hired 15 to <30 years before baseline, and hired within 15 years of baseline. For this stratified analysis, we simulated additional data to provide sufficient power. The simulation conditions remained identical except that all 5000 subjects in each cohort were hired between 1900 and 1949, making anyone still at work in 1950 a prevalent hire.

RESULTS

First we address the reproducibility of the simulation for the exposure response for work duration (β_1) in incident hires only, without variability in susceptibility ($\beta_3 = 0$, $\beta_4 = 0$) (Table 1). The mean and standard deviation (SD) of the $\hat{\beta}_1^*$ estimates across the 500 simulated cohorts were good estimates of the true parameters for a null ($\hat{\beta}_1^*=0.02$ [SD = 0.03]) and a positive association ($\hat{\beta}_1^*=0.51$[SD = 0.05]).

As also shown in Table 1, we considered the impact of susceptibility on $\hat{\beta}_1^*$ for work duration among incident hires. When only disease susceptibility was present ($\beta_3 = 0.5$, $\beta_4 = 0$), the estimated mean exposure-response parameter, $\hat{\beta}_1^*$, reproduced both the underlying null and positive associations, regardless of the percentage of susceptible subjects included in the population. Introducing only disease susceptibility, however, did increase the background rate of disease, as observed in the increased number of cases.

Under the null, when susceptibility to the effect of exposure alone ($\beta_3 = 0$, $\beta_4 = 0.5$) or in combination with susceptibility to disease ($\beta_3 = \beta_4 = 0.5$) was present, $\hat{\beta}_1^*$ was elevated above the true value of β_1 (Table 1) and the estimate increased as more susceptible subjects were included, (eg, null, $\beta_3 = \beta_4 = 0.5$; $\hat{\beta}_1^*=0.10$ [SD = 0.03] with 25% of subjects susceptible and $\hat{\beta}_1^*=0.18$ [SD = 0.03] with 50% of subjects susceptible). Note that $\hat{\beta}_1^*$ is a weighted average of 2 types of subjects: nonsusceptible subjects, whose association between work duration and disease mortality is simulated by β_1 only, and susceptible subjects, whose association is simulated by β_1 and β_4. When the percentage of susceptible subjects increases, the number of cases increases. This same pattern was observed for a positive association.

Moving into comparisons between incident and prevalent hires, we restrict presentation to scenarios without susceptibility ($\beta_3 = \beta_4 = 0$) and with susceptibility to both disease and to the effect of exposure ($\beta_3 = \beta_4 = 0.5$). We do not present scenarios with disease

susceptibility only ($\beta_3 = 0.5$, $\beta_4 = 0$) because conclusions about prevalent hires were the same as the simulation without susceptibility. Likewise, we do not present scenarios with susceptibility only to the effect of exposure ($\beta_3 = 0$, $\beta_4 = 0.5$), as conclusions were the same as with susceptibility to both disease and to the effect of exposure.

In Table 2, mean cohort characteristics for incident and prevalent hires are compared, as generated under the null without susceptibility. The size of the simulated cohorts was always 5000 subjects for incident hires, whereas the size of the subsets of prevalent hires varied, with approximately 1200 subjects, on average. About one-third of incident and prevalent hires became cases, on average, and case age at death was somewhat older for prevalent hires. Mean length of follow-up and work duration were longer in prevalent than incident hires. By definition, incident hires included all workers hired during follow-up. On the other hand, prevalent hires had already worked 16.2 (SD = 0.3) years, on average, before follow-up started.

Next, we compared mean $\hat{\beta}_1^*$ for incident and prevalent hires under various simulation conditions (Table 3). Under the null without susceptibility, the exposure-response estimates were not different. In the presence of susceptibility (either 25% or 50% of subjects susceptible), differences emerged, with $\hat{\beta}_1^*$ lower in prevalent hires than incident hires. Similar patterns were observed for a positive association; no differences between incident and prevalent hires in the absence of susceptibility, and a lower $\hat{\beta}_1^*$ for prevalent hires than incident hires in the presence of susceptibility.

Prevalent hires contain subjects who worked up to 50 years before the start of follow-up. To further examine bias from prevalent hires, they were stratified by length of time worked prior to follow-up: hired <15, 15 to <30, and ≥30 years before baseline. In Figure 1, we observe that under the null, without susceptibility, $\hat{\beta}_1^*$ returned the null value, regardless of how far in advance of baseline they had started working ($\hat{\beta}_1^*=0.002$ [SD = 0.10], 0.001 [0.27], and −0.02 [1.06] for worked <15, 15 to <30, and ≥30 years before baseline, respectively). By contrast, when 25% of subjects were susceptible to the effect of exposure, a negative bias appeared, becoming more pronounced with increasing years worked before baseline (null, 25% susceptible: $\hat{\beta}_1^*=0.05$ [SD = 0.08], −0.08 [0.24], and −0.18 [0.98] for worked <15, 15 to <30, and ≥30 years prior to baseline, respectively). This negative bias was most evident when 50% of subjects were susceptible (null, 50% susceptible: $\hat{\beta}_1^*=0.12$ [SD = 0.08], −0.08 [0.21], and −0.28 [0.92] for worked <15, 15 to <30, and ≥30 years prior to baseline, respectively).

For a positive association without heterogeneous susceptibility, the simulated value ($\beta_1 = 0.5$) was reproduced regardless of the length of time prevalent hires had worked prior to follow-up ($\hat{\beta}_1^*=0.50$ [SD = 0.14], 0.53 [0.35], and 0.59 [1.23] for hired <15, 15 to <30, and ≥30 years before baseline, respectively) (Fig. 2). However, with heterogeneous susceptibility, $\hat{\beta}_1^*$s are increasingly pulled downward with more time between hire and baseline ($\hat{\beta}_1^*$ for 25% susceptible subjects = 0.51 [SD = 0.12], 0.32 [0.28], and 0.05 [1.11]

for hired <15, 15 to <30, and ≥30 years prior to baseline, respectively). In fact, when 50% of subjects were susceptible, the positive association was pulled below the null for those who had worked longest before baseline ($\hat{\beta}_1^*=0.58$ [SD = 0.10], 0.33 [0.24], and −0.12 [0.86] for worked <15, 15 to < 30, and ≥30 years prior to baseline, respectively).

To further examine the source of this downward bias when susceptibility was present, we evaluated the percent of susceptible subjects in the prevalent hires strata (Table 4). Under the null, with 25% susceptible subjects, the mean percentage of susceptible subjects among prevalent hires overall dropped to 21%, indicating that not all susceptible subjects made it into the follow-up period. Those hired more recently most closely resembled the original percent of susceptible subjects, whereas those hired longer ago had a much lower percent of susceptible subjects (eg, under the null, with 25% susceptible: 24.2%, 19.6%, and 12.0% of the subjects were susceptible, on average, among subjects hired <15, 15 to <30, and ≥30 years before baseline, respectively). This pattern of a decreasing percentage of susceptible subjects with greater time between hire and start of follow-up was evident for both null and positive associations with 25% and 50% susceptible subjects.

DISCUSSION

Employer records usually form the basis for defining occupational cohorts. Because employment records are generally maintained for all active workers, study populations typically include presently active workers as well as new employees hired during the subsequent follow-up period. Results from our simulation study are relevant to cohorts that comprise prevalent hires as a part or all of the study population.

Workers hired before the start of follow-up, ie, prevalent hires, can introduce downward bias in exposure-response estimates for null and positive associations. Moreover, this bias increases with time between hire and start of follow-up, driven by a change in the composition of surviving subjects over time; those more susceptible to the effect of exposure are less likely to remain at work until the later start of follow-up. Those who do remain at work are less diverse and comprise a larger proportion of nonsusceptibles than their counterparts who are no longer followed. This result, however, was observed only when we explicitly incorporated heterogeneity in susceptibility to the effect of exposure. Susceptibility to the effect of disease alone did not influence bias due to prevalent hires (data not shown). Because people do in fact vary in their biologic response to many exposures—due to genetic or epigenetic factors, early life events, and other unmeasured characteristics—the assumption of variable susceptibility is more plausible than homogeneous responsiveness.

Our simulation model for disease risk included independent terms for susceptibility to disease, susceptibility to the effect of exposure, and an interaction between disease susceptibility and exposure with a positive coefficient. This is only one of the possible scenarios. The epidemiology literature[20,21] describes many types of interactions. Although how the bias changes over time may differ from the scenarios presented here, we suspect that the conclusion would be the same: the exposure-response relationship changes over time because the pattern of heterogeneity in the population at risk changes over time.

In our simulation study, analysis restricted to recent prevalent hires (hired within 15 years of baseline) produced unbiased results. We expect this to hold for diseases with similar induction time from first exposure to diagnosis; the period would be shorter for diseases with a shorter induction time. A sensitivity analysis designed to vary the eligibility criteria (based on time hired prior to baseline) might provide information about the bounds of the bias due to the presence of left-truncated data.[11] For studies that include a mixture of prevalent and incident hires, we advise against controlling for hire status as a confounder because doing so would provide a pooled estimate of the average exposure response for the incident and prevalent hires. Instead, it would be preferable to directly address the bias induced by prevalent hires.

Next, we describe several analytic approaches that may be useful for addressing prevalent-hire bias. We describe them here but more direct evaluation of these methods is needed. The first involves partitioning exposure for prevalent hires into 2 time windows: before the start of follow-up (a baseline constant) and during the follow-up period (time-varying) (J Robins, personal communication, 2008). This approach models disease as a function of exposure accumulated during the follow-up period, adjusting for cumulative exposure prior to start of follow-up (coded as zero for incident hires). However, the same cumulative exposure can be reached through various combinations of duration and intensity. As we have seen in this analysis, bias from prevalent hires is greater with longer duration employed before baseline. Therefore, partitioning exposure alone may not be sufficient, but is perhaps best done in conjunction with a method that addresses the change in composition of subjects over time.

There are potentially several ways of addressing the surviving subjects over time. First, if the proportional hazards assumption is not satisfied, an extended Cox model may be used to handle time-dependent effects.[22] This method has been recommended for left truncation in studies of cancer survivors where cases were interviewed some time after diagnosis and then followed.[7] Alternatively, prevalent hires could be handled with delayed entry using time since hire as the study timeline.[23–25] This approach involves a likelihood function conditional on having stayed at work until a particular time. Incident hires contribute person-time to risk sets beginning at hire. For each prevalent hire, the starting time is the time between hire and start of follow-up (eg, delayed entry). Thus, prevalent hires will begin to contribute person-time to the same risk set as incident hires who survived until the same amount of time after hire. Finally, because the composition of subjects, and thus the association of interest, may change over time, a single summary hazard ratio may be inappropriate. An alternative is to present a series of hazard ratios by increasing categories of time since hire. As recently described by Hernán,[26] standard structural accelerated failure time models for survival ratios or the periodic estimation of association over the follow-up period may be preferable to a single summary hazard ratio.

Studies comprising only incident hires (inception cohorts) also have limitations. First, the survivor issue driving prevalent-hire bias operates in incident hires as well. That is, as time in follow-up increases, more susceptible subjects will be more likely to experience the disease and experience it earlier than the less susceptible. Thus, the methods described earlier might well be applied in long-term studies of incident hires as well. Moreover, restricting a study to incident hires may reduce statistical power due to a decrease in the

range of exposure or a reduction in the age range at baseline, with fewer subjects eligible for age-related diseases. In addition, the length of follow-up may be insufficient for a disease with longer latency if all subjects enter follow-up when hired.[10] Although a prospective study of an inception cohort may sometimes be feasible, a retrospective inception cohort of all workers hired after the start of follow-up may be more practical.

Both the healthy worker survivor effect and depletion of susceptible individuals have been offered as explanations for the frequently observed attenuation of exposure response at the highest levels of workplace exposures.[27] Recall that healthy-worker-survivor-effect bias occurs if less healthy workers transfer to jobs with lower exposure, take time off, or leave work altogether prior to the event of interest (eg, death). Thus, susceptibility is also inherent in the healthy worker survivor effect. In fact, with the healthy worker survivor effect, workers with greater susceptibility to the exposure effect deliberately intervene in their exposure trajectories in response to an early health event or symptom. Thus it can be considered a problem of a time-varying confounder (eg, underlying health status), also affected by prior exposure. Although standard epidemiologic risk models cannot address a confounder on the causal pathway, causal models, in particular, g-estimation of structural nested accelerated failure time models, can be applied to such situations to address healthy-worker-survivor-effect bias.[28] It is important to note that the prevalent-hire bias we observed in this study was operating in the absence of the healthy worker survivor effect; we did not simulate workers with exposure-related susceptibility taking time off work, thereby reducing their employment duration. Moreover, left truncation in the absence of the healthy worker survivor effect does not involve workers changing their subsequent exposure on the basis of their health status, ie, the bias is not caused by a time-varying confounder on the causal pathway. Thus, although susceptibility is involved in both biases, the mechanisms differ, and alternative methods (such as those described earlier) are needed to reduce bias from prevalent hires alone.

In conclusion, this simulation has shown that bias is induced by including prevalent hires with heterogeneity in their susceptibility to the effect of exposure. Moreover, the bias increased with time employed prior to start of follow-up, lending support to the presumption that subjects who survive longer are different from those who do not. This lack of exchangeability is magnified by including prevalent hires, but also exists within an incident-hire cohort with long-term follow-up, and can cause bias even in the absence of the healthy worker survivor effect. Descriptions of the published cohorts should be sufficient to determine the mixture of prevalent and incident hires. In addition, alternative analytic strategies should be considered to address the potential bias due to differential survival.

Acknowledgments

Supported by the Centers for Disease Control—National Institute of Occupational Safety and Health grant K01OH009390 (to K.M.A.). These contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute or the National Institute of Occupational Safety and Health.

References

1. Szklo M. Design and conduct of epidemiologic studies. Prev Med. 1987; 16:142–149. [PubMed: 3495793]
2. Brookmeyer R, Gail MH, Polk BF. The prevalent cohort study and the acquired immunodeficiency syndrome. Am J Epidemiol. 1987; 126:14–24. [PubMed: 3647718]
3. Brookmeyer R, Gail MH. Biases in prevalent cohorts. Biometrics. 1987; 43:739–749. [PubMed: 3427161]
4. Cole SR, Li R, Anastos K, et al. Accounting for leadtime in cohort studies: evaluating when to initiate HIV therapies. Stat Med. 2004; 23:3351–3363. [PubMed: 15493031]
5. Howards PP, Hertz-Picciotto I, Poole C. Conditions for bias from differential left truncation. Am J Epidemiol. 2007; 165:444–452. [PubMed: 17150983]
6. Hernàn MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol. 2002; 155:176–184. [PubMed: 11790682]
7. Azzato EM, Greenberg D, Shah M, et al. Prevalent cases in observational studies of cancer survival: do they bias hazard ratio estimates? Br J Cancer. 2009; 100:1806–1811. [PubMed: 19401693]
8. Koskela RS, Jarvinen E, Kolari PJ. Effect of cohort definition and follow-up length on occupational mortality rates. Scand J Work Environ Health. 1984; 10:311–316. [PubMed: 6335294]
9. Weiss W. Heterogeneity in historical cohort studies: a source of bias in assessing lung cancer risk. J Occup Med. 1983; 25:290–294. [PubMed: 6602212]
10. Checkoway, H.; Pearce, N.; Kriebel, D. Monographs in Epidemiology and Biostatistics. 2. New York: Oxford University Press; 2004. Research Methods in Occupational Epidemiology.
11. Applebaum KM, Malloy EJ, Eisen EA. Reducing healthy worker survivor bias by restricting date of hire in a cohort study of Vermont granite workers. Occup Environ Med. 2007; 64:681–687. [PubMed: 17449560]
12. Richardson D, Wing S, Steenland K, McKelvey W. Time-related aspects of the healthy worker survivor effect. Ann Epidemiol. 2004; 14:633–639. [PubMed: 15380793]
13. Steenland K, Stayner L. The importance of employment status in occupational cohort mortality studies. Epidemiology. 1991; 2:418–421. [PubMed: 1790193]
14. Delzell E, Monson RR. Mortality among rubber workers. Part IV: General mortality patterns. J Occup Med. 1981; 23:850–856.
15. Gilbert ES. Some confounding factors in the study of mortality and occupational exposures. Am J Epidemiol. 1982; 116:177–188. [PubMed: 7102651]
16. Weiss W. Lung cancer due to chloromethyl ethers: bias in cohort definition. J Occup Med. 1989; 31:102–105. [PubMed: 2785169]
17. Stayner L, Smith R, Schnorr T, Lemen R, Thun M. Lung cancer. Ann Epidemiol. 1993; 3:114–116. [PubMed: 8287147]
18. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol. 1986; 15:413–419. [PubMed: 3771081]
19. D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. Stat Med. 1990; 9:1501–1515. [PubMed: 2281238]
20. Khoury MJ, Adams MJ Jr, Flanders WD. An epidemiologic approach to ecogenetics. Am J Hum Genet. 1988; 42:89–95. [PubMed: 3337114]
21. Rothman, KJ.; Greenland, S.; Lash, TL. Modern Epidemiology. 3. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.
22. Kleinbaum, DG.; Klein, M. Statistics for Biology and Health. 2. New York: Springer; 2005. Survival Analysis: a Self-Learning Text.
23. Keiding, N. Survival and event history analysis. In: Andersen, PK.; Keiding, N., editors. Wiley Reference Series in Biostatistics. Chichester, West Sussex, England, Hoboken, NJ: Wiley; 2006. p. 173-178.

24. Klein, JP.; Moeschberger, ML. Survival Analysis: Techniques for Censored and Truncated Data. Statistics for Biology and Health. 2. New York: Springer; 2003.

25. Brookmeyer, R. Biased sampling of cohorts. In: Armitage, P.; Colton, T., editors. Encyclopedia of Biostatistics. 2. Chichester, West Sussex, England;Hoboken, NJ: John Wiley; 2005. p. 427-439.

26. Hernán MA. The hazards of hazard ratios. Epidemiology. 2010; 21:13–15. [PubMed: 20010207]

27. Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I. Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. Scand J Work Environ Health. 2003; 29:317–324. [PubMed: 12934726]

28. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period - application to control of the healthy worker survivor effect. Math Mod. 1986; 7:1393–1512.

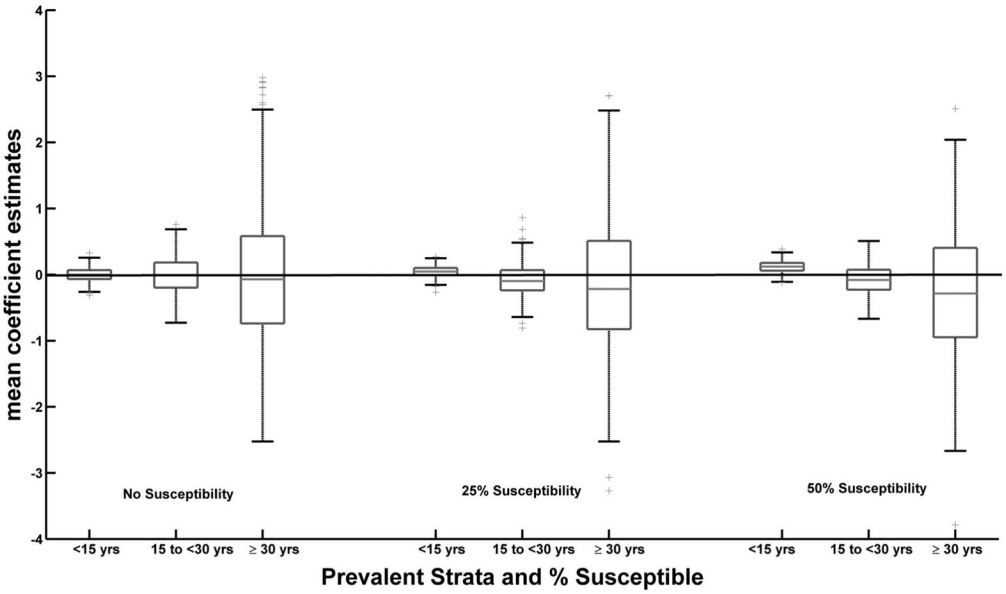

FIGURE 1.

Box plots of $\hat{\beta}_1^*$ estimates across 500 simulated cohorts under the null for prevalent hires, subjects who worked up to 50 years before the start of follow-up and still working when follow-up began. Prevalent hires were divided into strata: hired <15, 15 to <30, and ≥30 years before start of follow-up. Horizontal line indicates the simulated null value. Simulating model: logit $P = \beta_0 + = \beta_1$(ln work duration) + β_2 (ln age) = β_3 (disease susceptible) = β_4 (disease susceptible*ln work duration), with values set to $\beta_0 = -5$, $\beta_1 = 0$, $\beta_2 = 2$, and $\beta_3 = \beta_4 = 0$ for no susceptibility or = 0.5 with variable susceptibility (and either 25% or 50% of cohort susceptible). Data analysis model: $\log(\mathrm{HR}) = \hat{\beta}_1^*(\ln \text{ work duration})$ with age as the timeline.

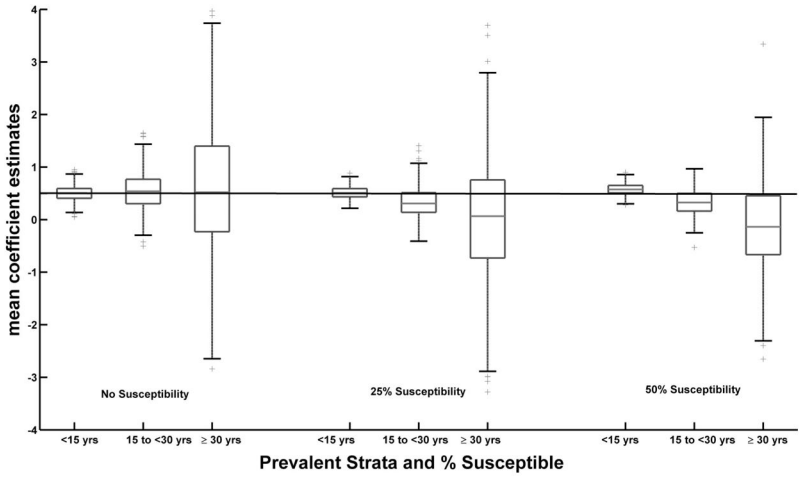

FIGURE 2.

Box plots of $\hat{\beta}_1^*$ estimates across 500 simulated cohorts for a positive association in prevalent hires, subjects who worked up to 50 years before the start of follow-up and still working when follow-up began. Prevalent hires were divided into strata: hired < 15, 15 to <30, and ≥30 years before start of follow-up. Horizontal line indicates the simulated positive value of $\beta_1 = 0.5$. Simulating model: logit $P = \beta_0 - \beta_1$(ln work duration) + β_2(ln age) + β_3(disease susceptible) + β_4(disease susceptible*ln work duration), with values set to $\beta_0 = -7$, $\beta_1 = 0.5$, $\beta_2 = 1$, and $\beta_3 = \beta_4 = 0$ for no susceptibility or = 0.5 with variable susceptibility (and either 25% or 50% of cohort susceptible). Data analysis model: $\log(\mathrm{HR}) = \hat{\beta}_1^*(\ln \text{ work duration})$ with age as the timeline.

TABLE 1

Mean $\hat{\beta}_1^*$ Estimate and Number of Cases for 500 Simulations of Incident Hire Cohorts With and Without Variability in Susceptibility Types

Simulation Conditions[a]				Mean Estimated Slope[b] and No. Cases	
ln work Duration	Susceptibility to Disease	Susceptibility to Effect of Exposure	% Subjects Susceptible	$\hat{\beta}_1^*$ (SD)	No. Cases (%)
Null[c] ($\beta_1 = 0$)	$\beta_3 = 0$	$\beta_4 = 0$	0	0.02 (0.03)	1714.8 (34.3)
	$\beta_3 = 0.5$	$\beta_4 = 0$	25	0.02 (0.03)	1896.3 (37.9)
			50	0.02 (0.03)	2077.5 (41.6)
	$\beta_3 = 0$	$\beta_4 = 0.5$	25	0.13 (0.03)	2261.1 (45.2)
			50	0.23 (0.03)	2804.2 (56.1)
	$\beta_3 = 0.5$	$\beta_4 = 0.5$	25	0.10 (0.03)	2386.1 (47.7)
			50	0.18 (0.03)	3059.8 (61.2)
Positive[d] ($\beta_1 = 0.5$)	$\beta_3 = 0$	$\beta_4 = 0$	0	0.51 (0.05)	1013.9 (20.3)
	$\beta_3 = 0.5$	$\beta_4 = 0$	25	0.50 (0.05)	1142.6 (22.9)
			50	0.50 (0.05)	1273.8 (25.5)
	$\beta_3 = 0$	$\beta_4 = 0.5$	25	0.63 (0.05)	1507.7 (30.2)
			50	0.74 (0.04)	2001.4 (40.2)
	$\beta_3 = 0.5$	$\beta_4 = 0.5$	25	0.57 (0.04)	1675.9 (33.5)
			50	0.65 (0.04)	2338.3 (46.8)

[a] Simulating model: logit $P = \beta_0 + \beta_1$(ln work duration) + β_2(ln age) + β_3(disease susceptible) + β_4(disease susceptible*ln work duration).

[b] Data analysis model: $\log(\mathrm{HR}) = \hat{\beta}_1^*(\ln \text{ work duration})$ with age as the timeline.

[c] Simulation conditions for null models: $\beta_0 = -5$, $\beta_1 = 0$, $\beta_2 = 2$, and β_3 and $\beta_4 = 0$ or 0.5, as indicated; with either 0%, 25%, or 50% of subjects susceptible.

[d] Simulation conditions for positive models: $\beta_0 = -7$, $\beta_1 = 0.5$, $\beta_2 = 1$, and β_3 and $\beta_4 = 0$ or 0.5, as indicated; with either 0%, 25%, or 50% of subjects susceptible.

TABLE 2

Comparison of Mean Cohort Characteristics for Incident and Prevalent Hires From 500 Simulated Cohorts Under the Null With No Variability in Susceptibility[a]

	Incident Hires[b]	Prevalent Hires[c]
Cohort size		
No. (SD)	5000 (0)	1257.8 (31.7)
Cases		
No. (%)	1714.8 (34.3)	414.2 (32.9)
Case age at death (years); mean (SD)	65.3 (0.4)	69.8 (0.8)
Length of follow-up (years); mean (SD)	48.5 (0.3)	52.8 (0.5)
Work duration (years); mean (SD)	25.2 (0.2)	32.3 (0.3)
Time worked before baseline (years); mean (SD)	0 (0)	16.2 (0.3)

[a] 500 simulations contained 5000 subjects each. Simulating model: logit $P = \beta_0 + \beta_1$ (ln work duration) + β_2 (ln age) + β_3 (disease susceptible) + β_4 (disease susceptible*ln work duration). Simulation conditions were set to $\beta_0 = -5$, $\beta_1 = 0$, $\beta_2 = 2$, $\beta_3 = \beta_4 = 0$ (models had no heterogeneity in susceptibility).

[b] Incident hire subjects started working 1900–2000 and study follow-up was 1900–2050 (eg, hired during follow-up).

[c] Prevalent hire subjects started working 1900–1949 and were still at work in 1950 when follow-up began and ran through 2050 (eg, hired prior to follow-up and still at work when follow-up began).

TABLE 3

Comparison of Incident Hires and Prevalent Hires: Mean Coefficient Estimation From 500 Simulations, With and Without Variability in Susceptibility of Subjects

Simulation Conditions[a]			
ln work Duration	**% Subjects Susceptible**	**Incident Hires[b] $\hat{\beta}_1^*$ (SD)**	**Prevalent Hire[c] $\hat{\beta}_1^*$ (SD)**
Null[d] ($\beta_1 = 0$)	0	0.02 (0.03)	−0.01 (0.13)
	25	0.10 (0.03)	0.03 (0.11)
	50	0.18 (0.03)	0.09 (0.10)
Positive[e] ($\beta_1 = 0.5$)	0	0.51 (0.05)	0.52 (0.19)
	25	0.57 (0.04)	0.49 (0.14)
	50	0.65 (0.04)	0.55 (0.11)

[a] 500 cohorts simulating with: logit $P = \beta_0 + \beta_1$ (ln work duration) + β_2(ln age) + β_3(disease susceptible) + β_4 (disease susceptible*ln work duration). Data analysis model: $\log(\mathrm{HR}) = \hat{\beta}_1^*(\ln \text{ work duration})$ with age as the timeline.

[b] Incident hires were hired during the follow-up period.

[c] Prevalent hires were hired up to 50 years prior to the start of follow-up and were still working when follow-up began.

[d] Simulation conditions for null models: $\beta_0 = -5$, $\beta_1 = 0$, $\beta_2 = 2$, and $\beta_3 = \beta_4 = 0$ for no susceptibility or = 0.5 with variable susceptibility (and there were either 25% or 50% of cohort susceptible).

[e] Simulation conditions for positive models: $\beta_0 = -7$, $\beta_1 = 0.5$, $\beta_2 = 1$, and $\beta_3 = \beta_4 = 0$ for no susceptibility or = 0.5 with variable susceptibility (and there were either 25% or 50% of cohort susceptible).

TABLE 4

Mean Percentage of Susceptible Subjects Among Prevalent Hires Only, by Years Hired Prior to Baseline

Simulation Conditions[a]		Mean Percentage of Susceptible Subjects by Years Hired Prior to Baseline			
ln work Duration	**% Subjects Susceptible**	**All**	**<15 Years**	**15 to <30 Years**	**≥30 Years**
Nullb ($\beta 1 = 0$)	25	21.0	24.2	19.6	12.0
	50	44.4	48.9	42.3	29.5
Positivec ($\beta 1 = 0.5$)	25	21.9	24.5	21.1	14.4
	50	45.7	49.4	44.5	33.4

[a] Following the simulation of 500 cohorts containing 5000 subjects each, the analysis was restricted to subjects meeting the definition of prevalent hire: hired before the start of follow-up and still at work when follow-up began (in this simulation, hired up to 50 years before baseline). Simulating model: logit $P = \beta_0 + \beta_1$(ln work duration) + β_2 (ln age) + β_3(disease susceptible) + β_4(disease susceptible × ln work duration); data analysis model: $\log(\mathrm{HR}) = \hat{\beta}_1^*(\ln \text{ work duration})$ with age as the timeline.

[b] Simulation conditions for the null models: $\beta_0 = -5$, $\beta_1 = 0$, $\beta_2 = 2$, $\beta_3 = 0.5$, $\beta_4 = 0.5$, with either 25% or 50% of subjects susceptible.

[c] Simulation conditions for the positive models: $\beta_0 = -7$, $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 0.5$, $\beta_4 = 0.5$, with either 25% or 50% of subjects susceptible.